Integration of Silicon Photonics into Electronic Processes

Jason Orcutt, Rajeev J. Ram, Vladimir Stojanovic
Massachusetts Institute of Technology

Photonics West (Silicon Photonics VIII)
4 February 2013

MIT-Micron DARPA POEM Team

Photonically **O**ptimized **E**mbedded **M**icroprocessors

MIT Primary Investigator

Vladimir Stojanović

Micron Project Leads

Gurtej Sandhu
Roy Meade

Micron Fabrication & Test Team

Efraim Megged
Zvi Sternberg
Ofer Tehar-Zahav
Reha Bafral
Yoel Shetrit
Harel Frish

MIT Post Processing

Hanqing Li

Photonic Design

Rajeev Ram
Milos Popovic *(U. C. Boulder)*
Michael Watts
Jeff Shainline *(U. C. Boulder)*
Karan Mehta
Erman Timurdgan

Link Circuit Design

Michael Georgas
Jonathan Leu
Ben Moss
Chen Sun

Architecture / Microprocessor Design

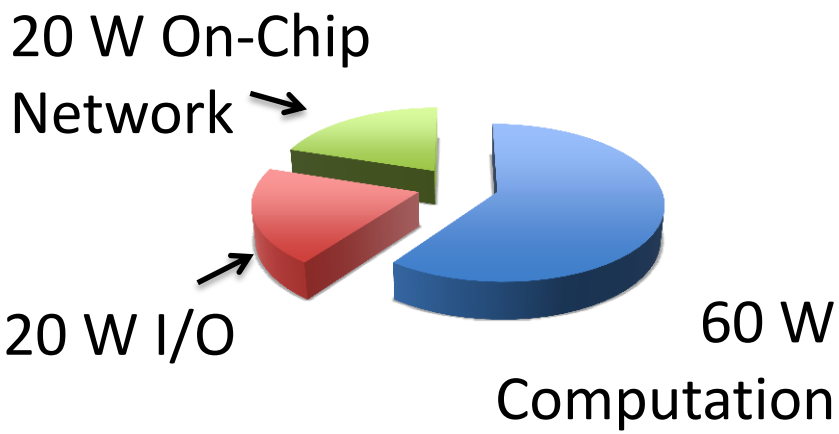
Krste Asanovic *(U. C. Berkeley)*
Yunsup Lee *(U. C. Berkeley)*

Work funded by DARPA under Dr. Jagdeep Shah

Core-to-Memory Scaling Issues

1 B per FLOP requires 48 Tb/s

Processors package-limited to
100 W power and 10,000 pins



Year	Technology	I/O Energy
'01-'03	DDR-333	257 pJ/bit
'05-'06	DDRII-667	121 pJ/bit
'09-'10	DDR3-1333	65 pJ/bit
'12-'14	DDR4-2667	39 pJ/bit
'13-'15	HMC	11 pJ/bit

Source: Micron Technology

10 pJ/bit would require 480 W

2015 UHPC Target = 10 pJ / FLOP

20 Gb/s requires 9600 I/O pins

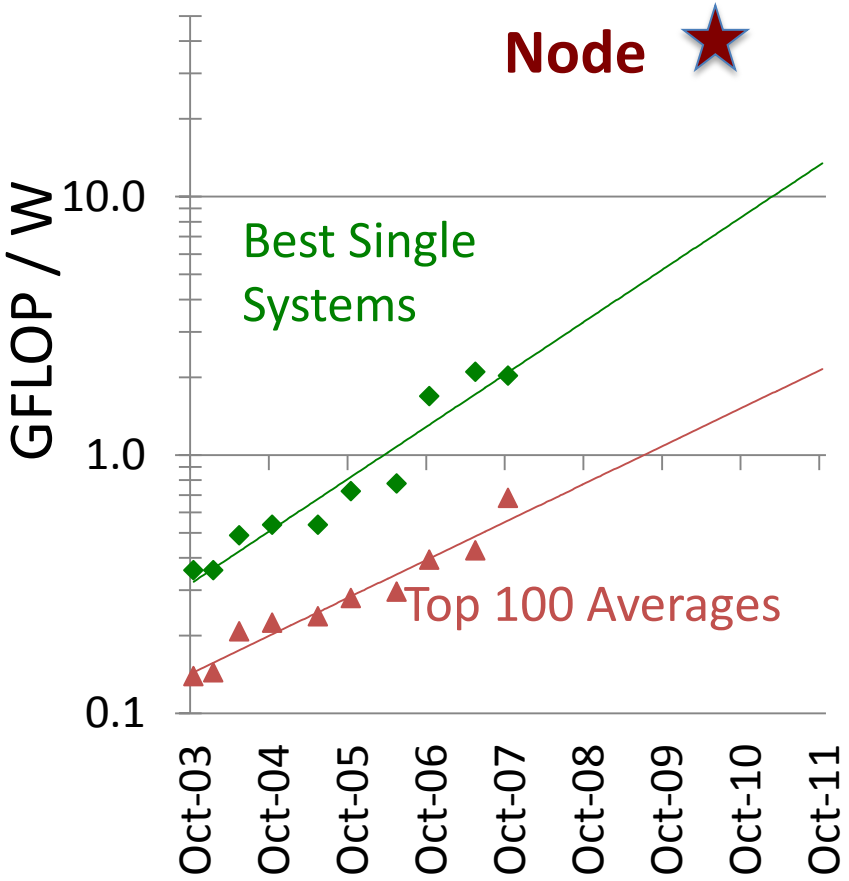
6 TFLOP chip possible

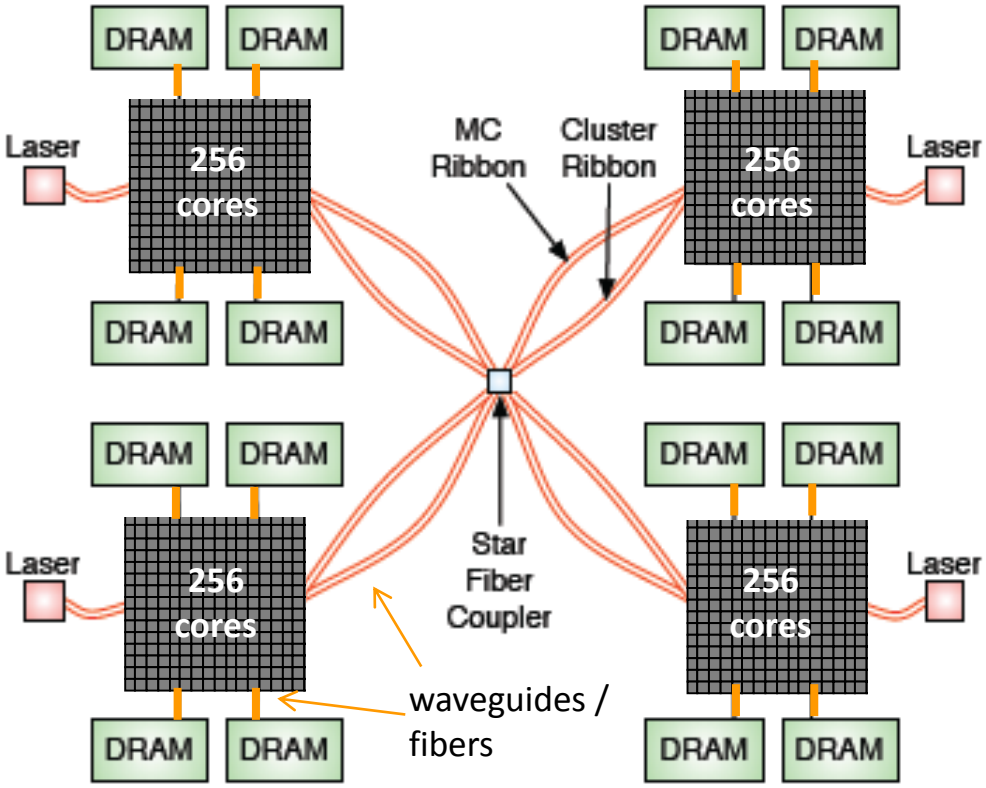
"Doubly Constrained" Problem

DARPA POEM Node Vision

S. Beamer et al., ISCA 2011

Green500 Supercomputer Scaling



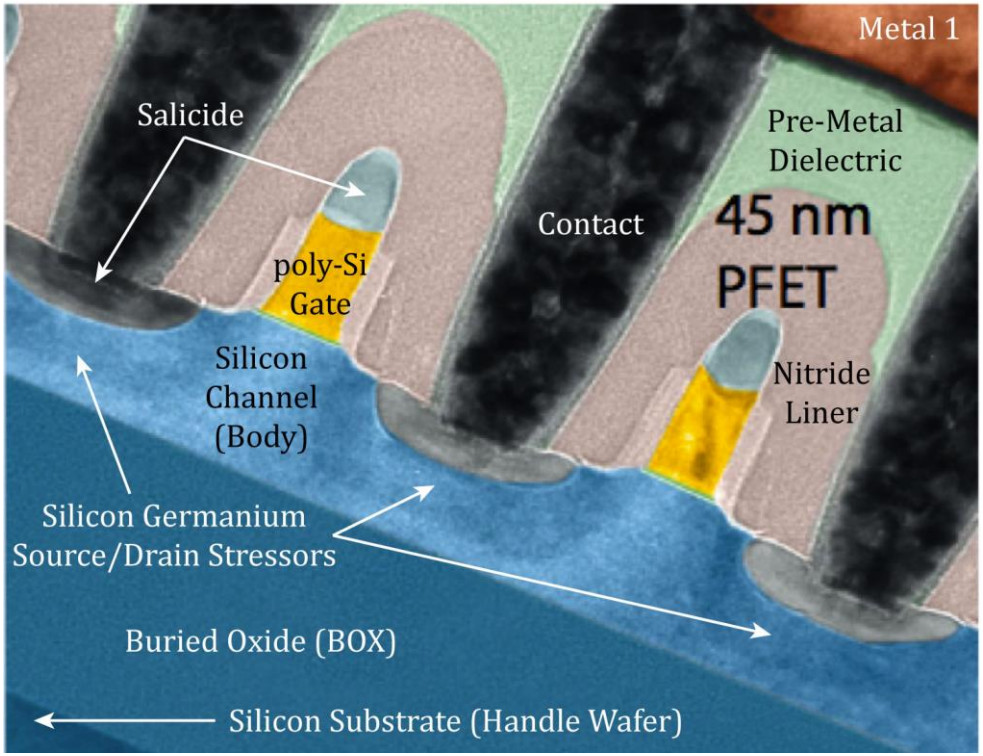


Computation	10 TFLOP
Total Power	285 W

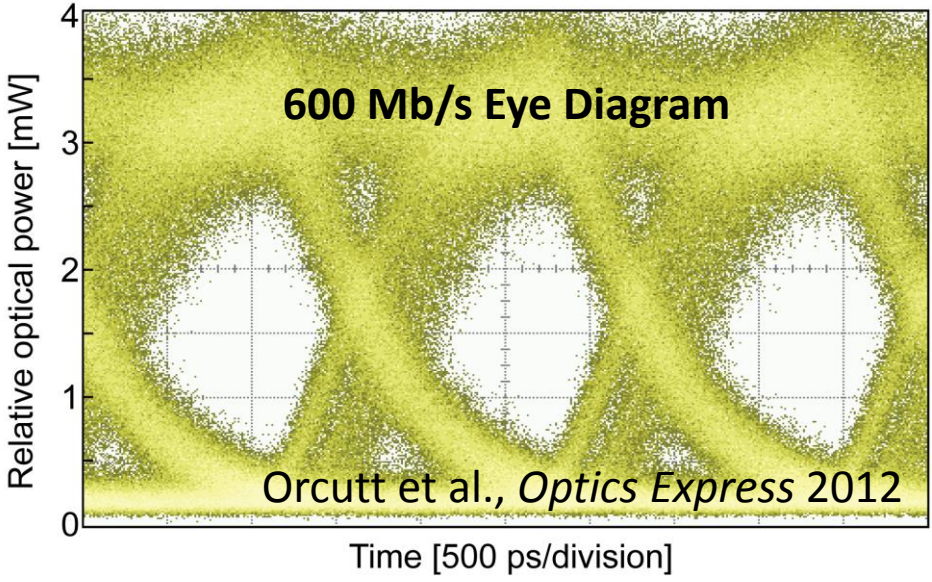
Wall-Plug Energy-Efficiency

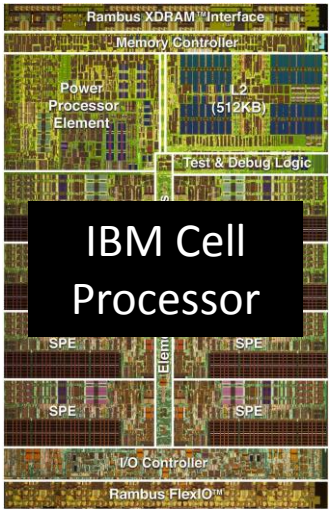
35 GFLOP / W

IBM 45nm CMOS Process (12SOI)

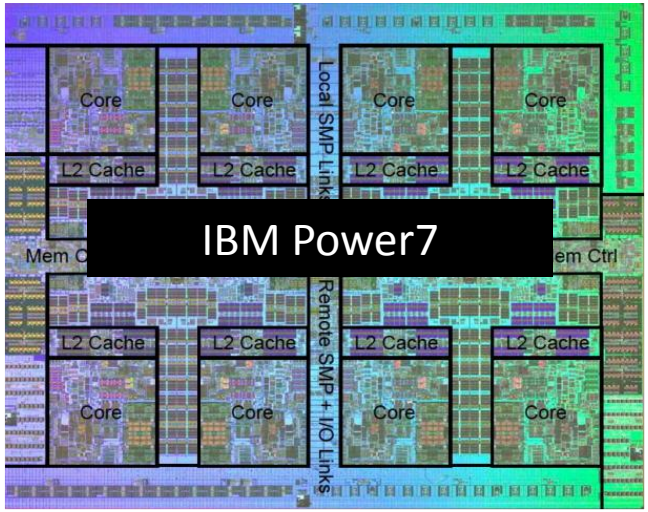


Available as a multi-project electronics process through the Trusted Access Program Office (TAPO)





IBM Cell Processor

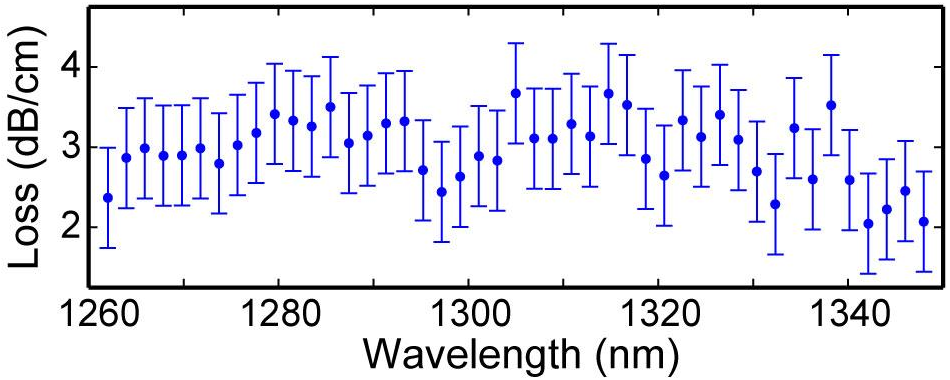


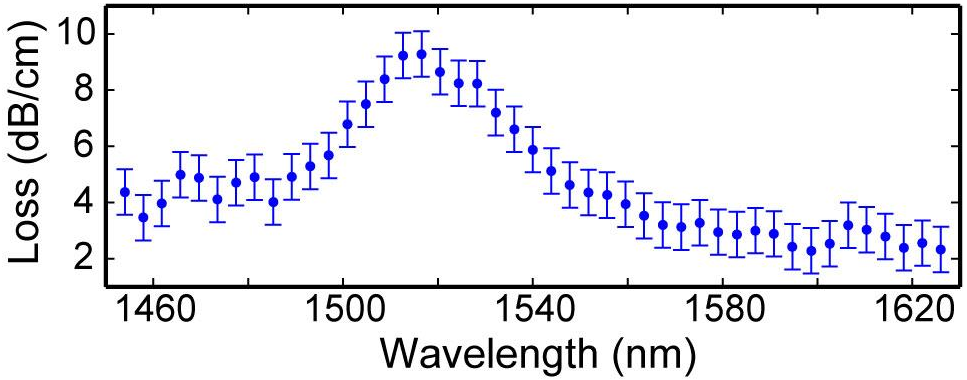
IBM Power7

Fully monolithic silicon photonic transmitter in a zero-change 45nm CMOS process demonstrated 2012

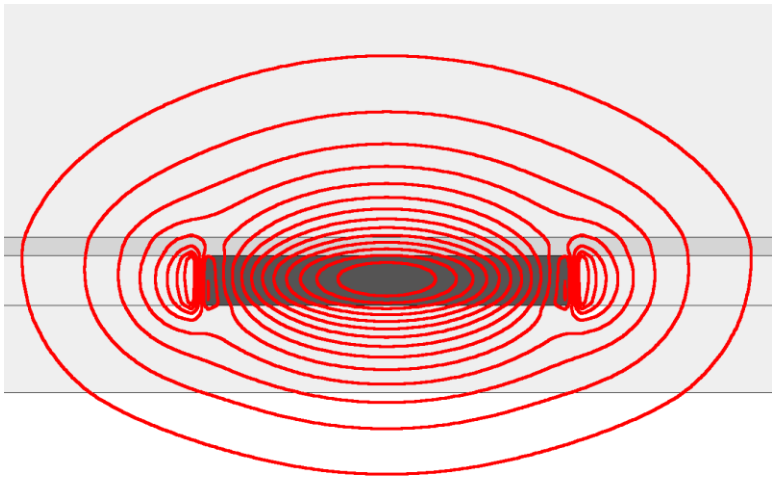
(2.5 Gb/s reported at ISSCC 2013)

High Performance Zero-Change Si Photonics



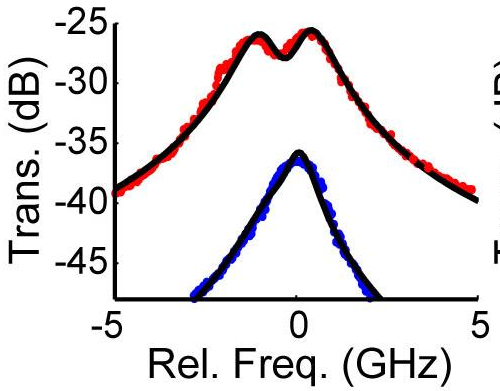


E-Field Contours of Integrated Mode

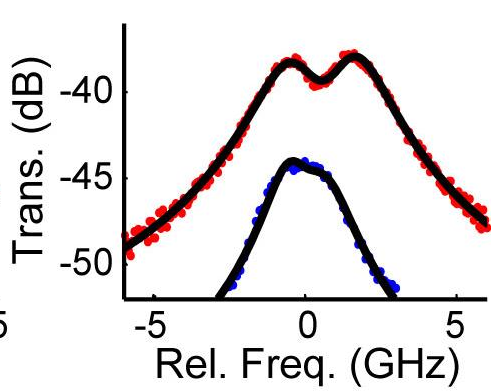


effective area of 0.16 μm²

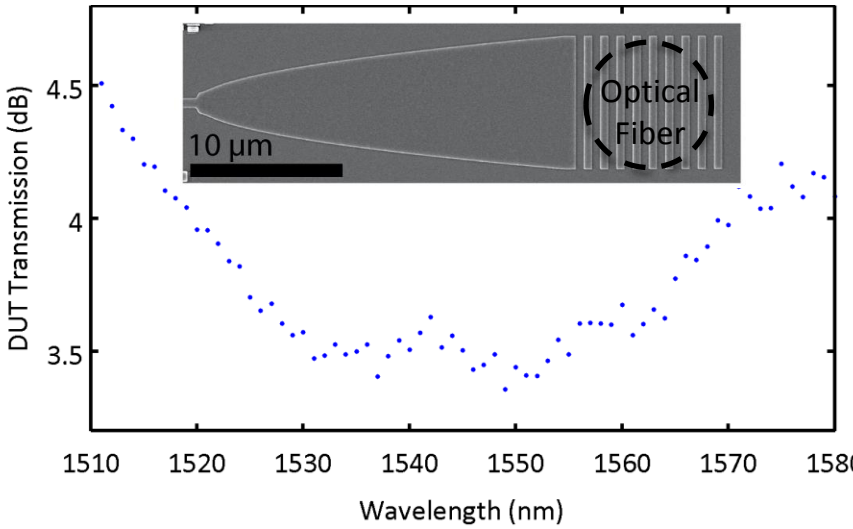
Q=227,000 @ 1.28μm



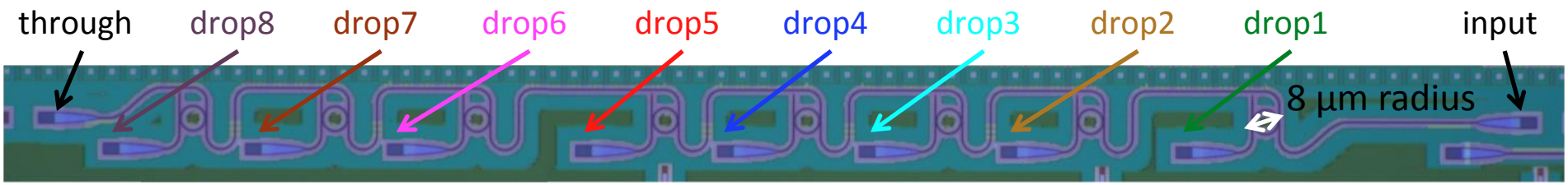
Q=112,000 @ 1.55μm

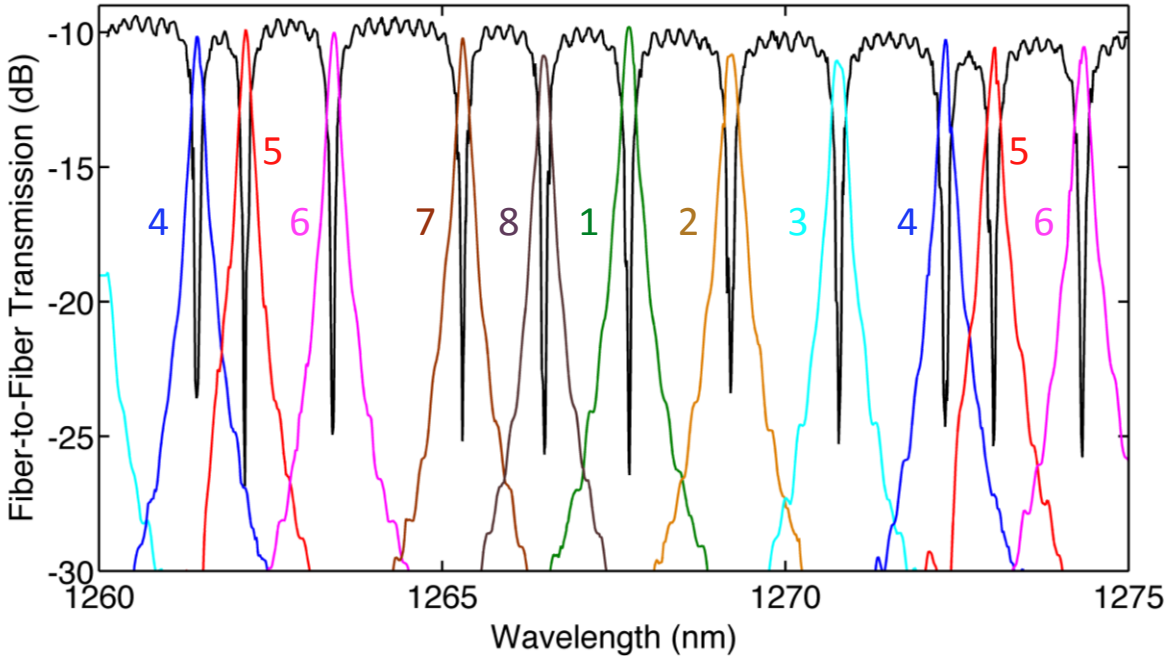


3.5 dB Bi-Directional Grating Coupler

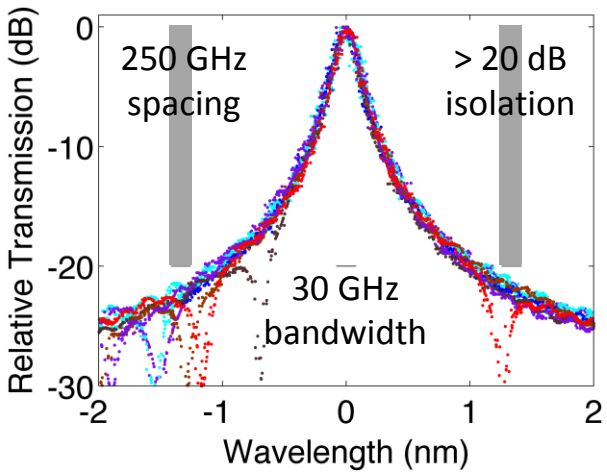


WDM Filter Bank





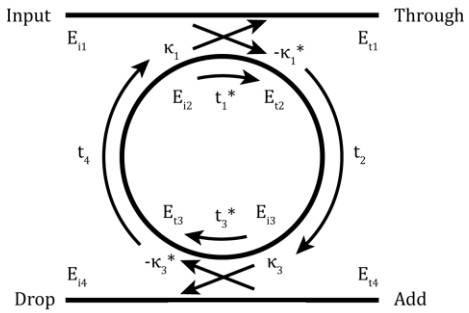
Superimposed Drop Ports

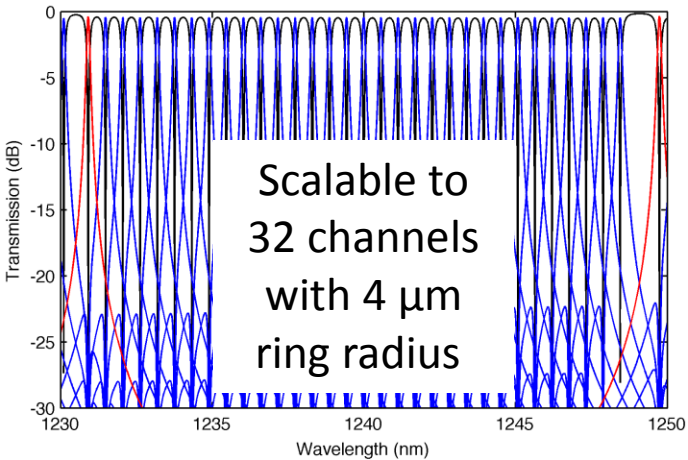


Drop loss is ~0.5 dB

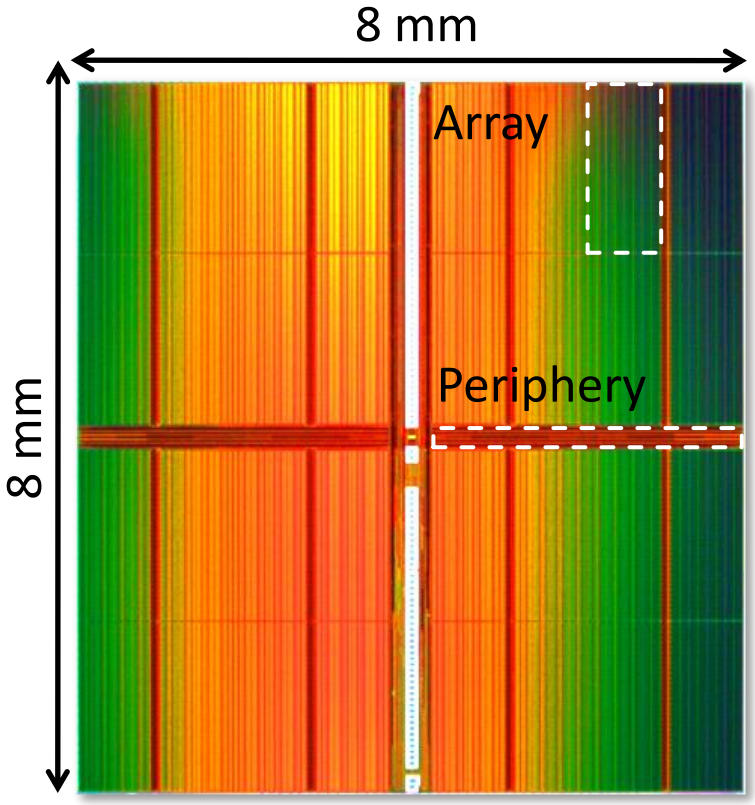
Cross talk <20 dB

Good relative alignment





Micron Photonic Memory Integration

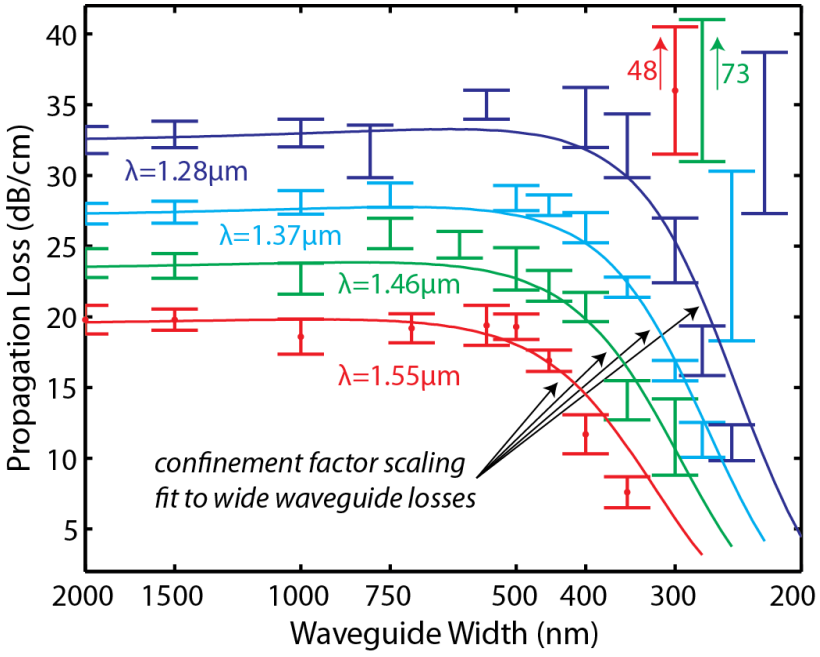


Process may be co-optimized for electronic and photonic device performance

Cost per transistor of less than 45 nanocents (n¢)

Polysilicon waveguide loss of 6.4 dB/cm)

2 Gb DDR3-1333
Die cost 90¢



Orcutt et al., *Optics Express* 2011

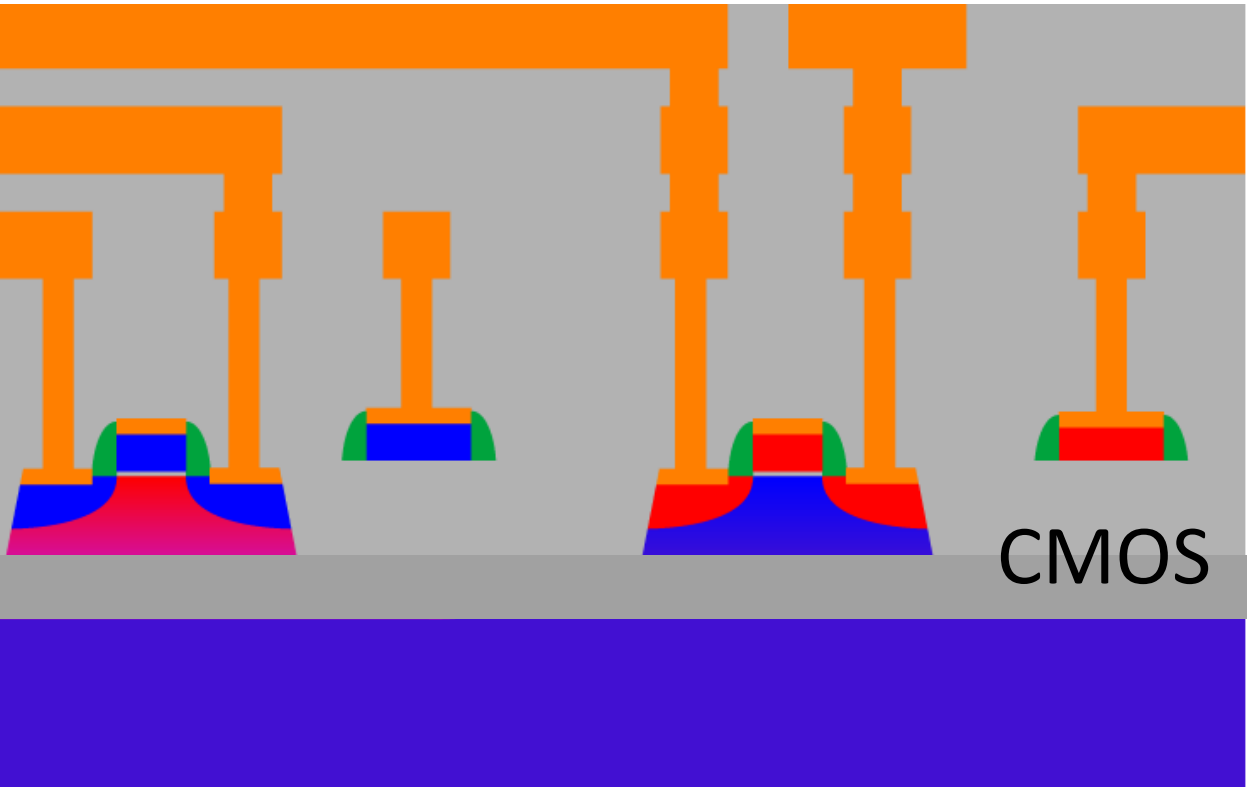
Monolithic Front-End Photonics

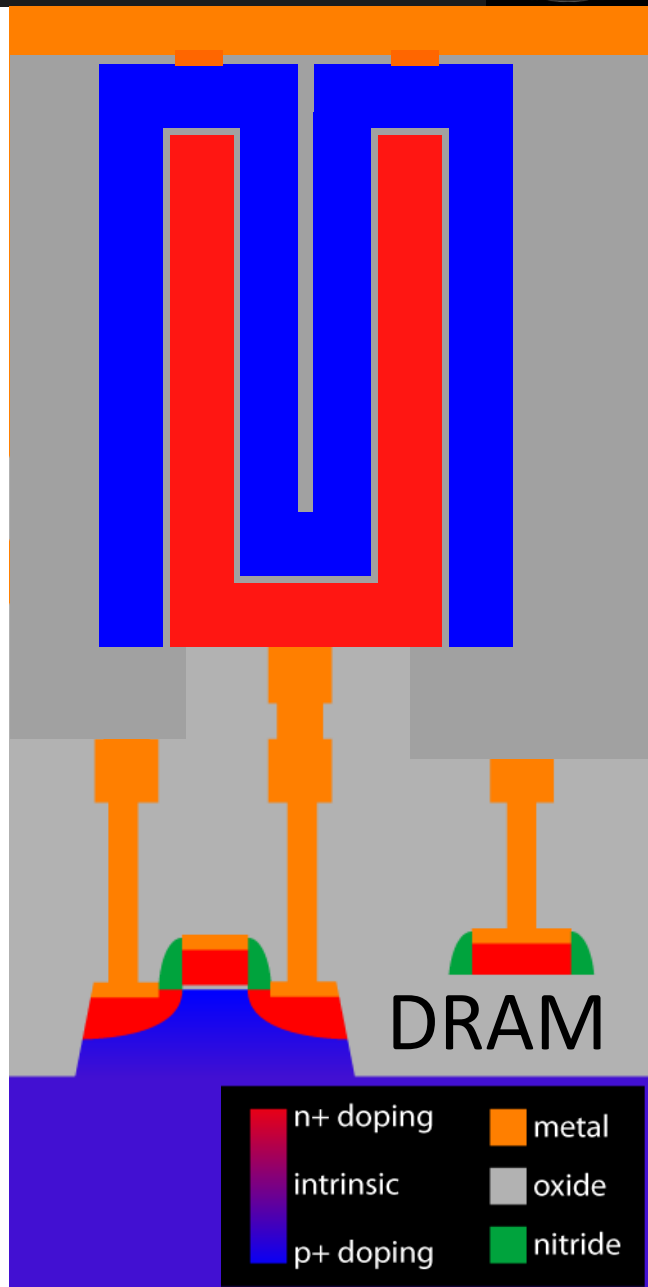
1. Process Integration
2. Area
3. Energy Efficiency
4. Alternative Schemes

Si Electronics Manufacturing Cross-Sections

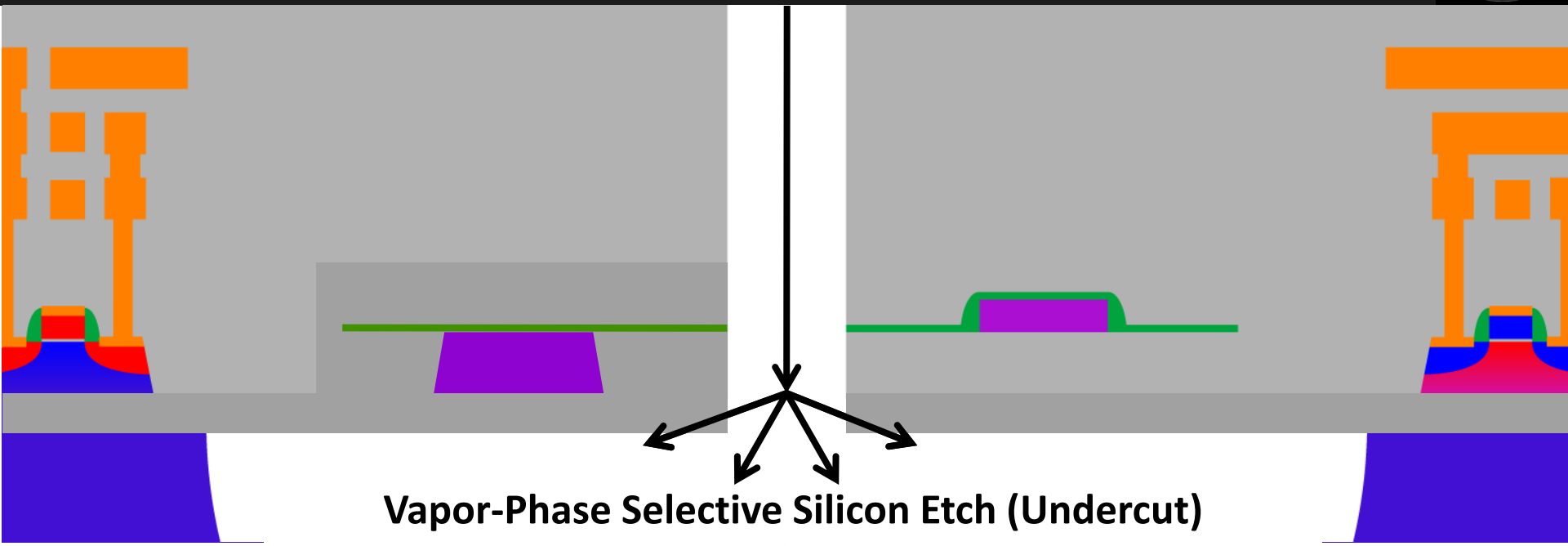
Standard processes include patternable silicon or poly-Si layers, but…

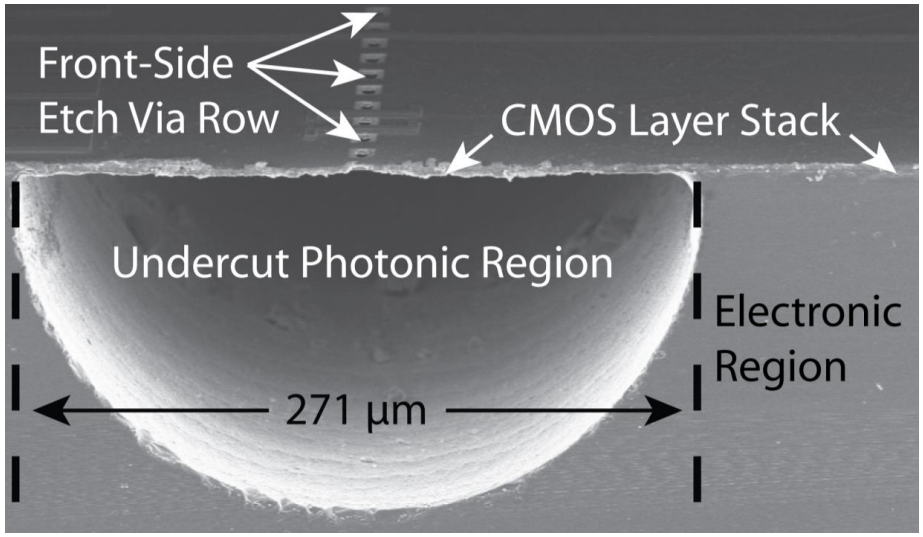
…these layers are heavily doped, metalized, surrounded by metal and in close proximity to a high-index substrate by default.





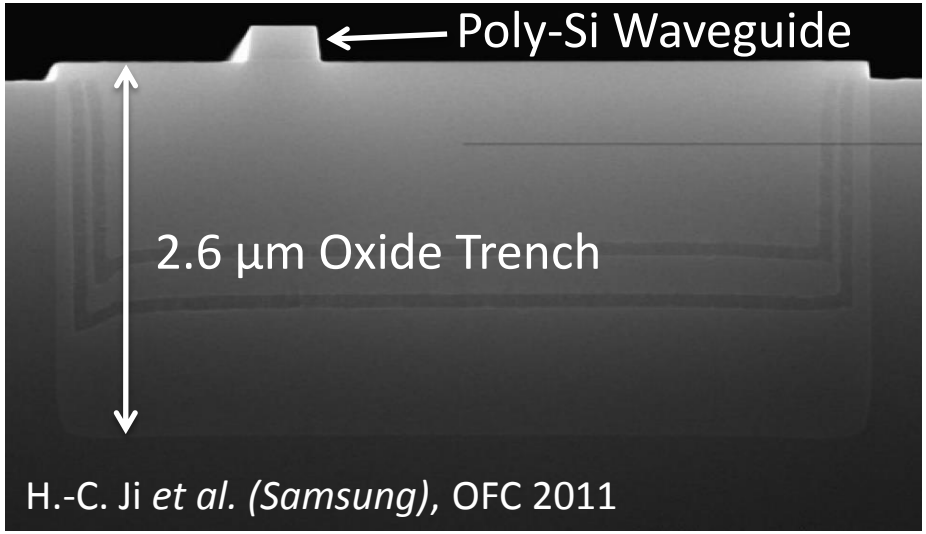
Waveguide Formation





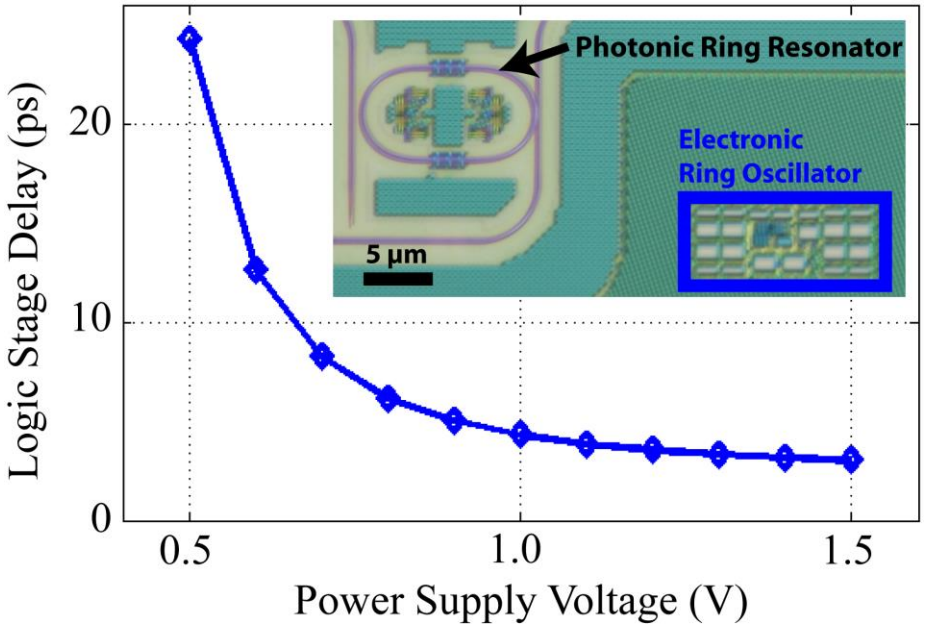
Orcutt *et al.*, OE 2011

Deep Oxide Trench Integration



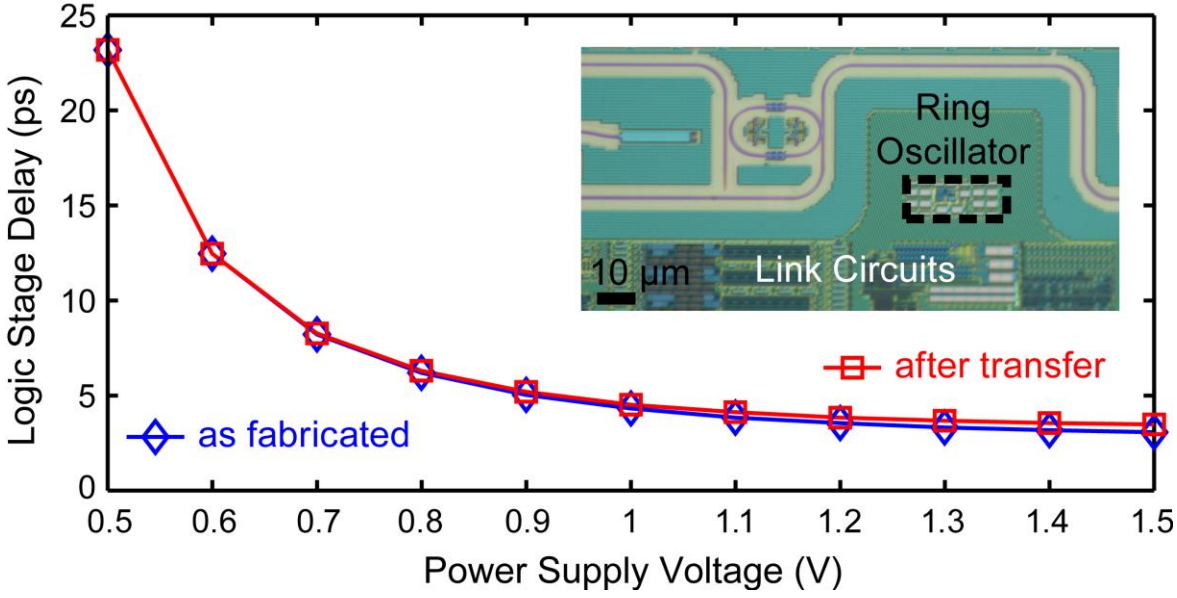
H.-C. Ji *et al. (Samsung)*, OFC 2011

Transistor Performance Verification



Included transistor test structures proximate to the photonic regions match performance targets.

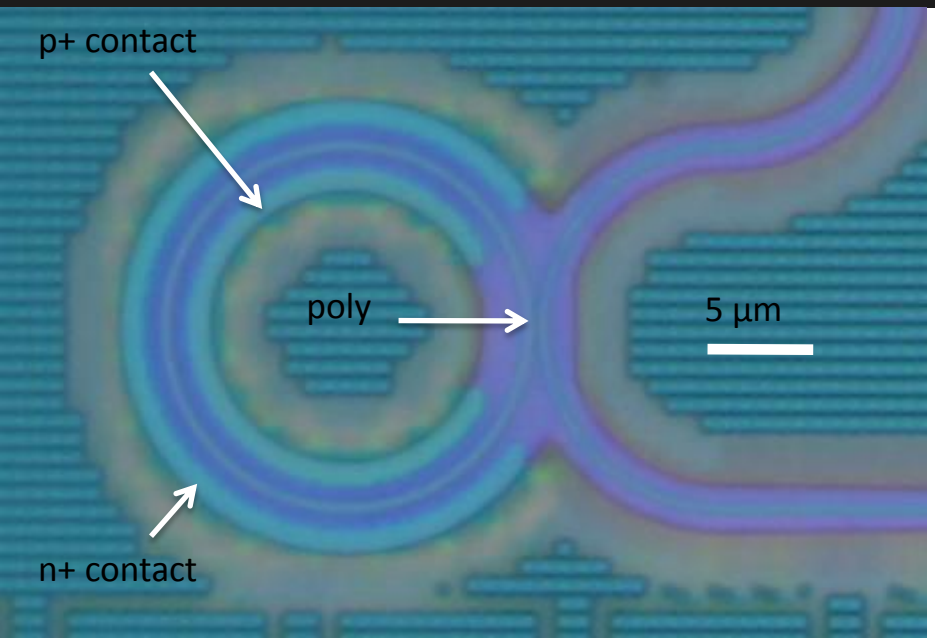
3-million transistor digital backend fully functional.



Less than 5% deviation in transistor characteristics observed after transfer.

Localized substrate removal technology would allow for unmodified local transistor environment.

Optical Modulator Integration



Integration into CMOS frontend provides access to many doping and contact steps

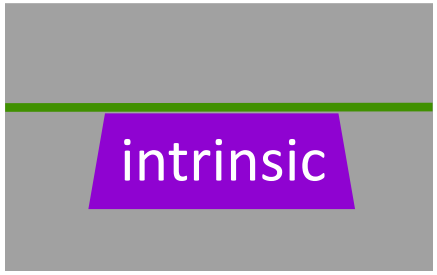
Opto-electronic modulation through carrier-injection into the optical mode

Access to transistor well doping levels enables depletion modulator functionality even in zero-change SOI-CMOS

Bulk CMOS / memory processes likely require custom implant levels

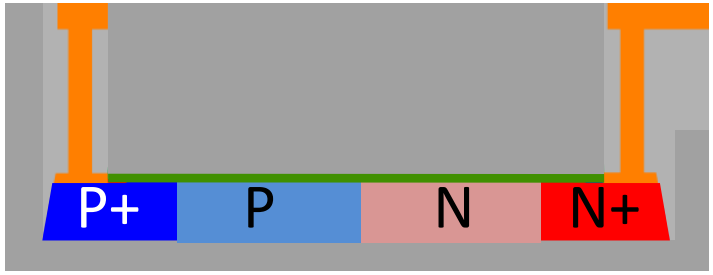
Doping Level Control

Waveguide

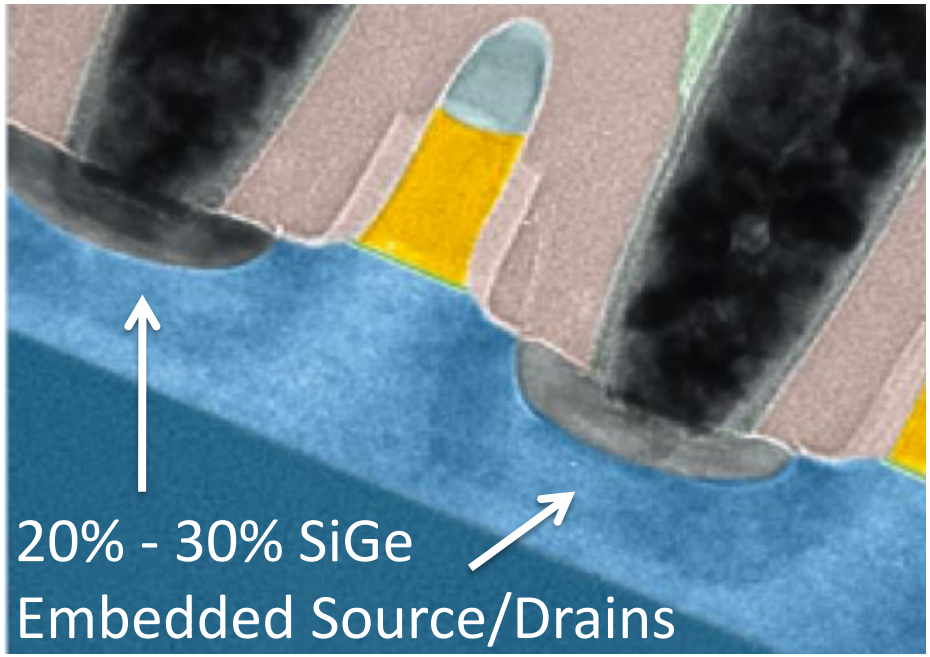


Injection Modulator

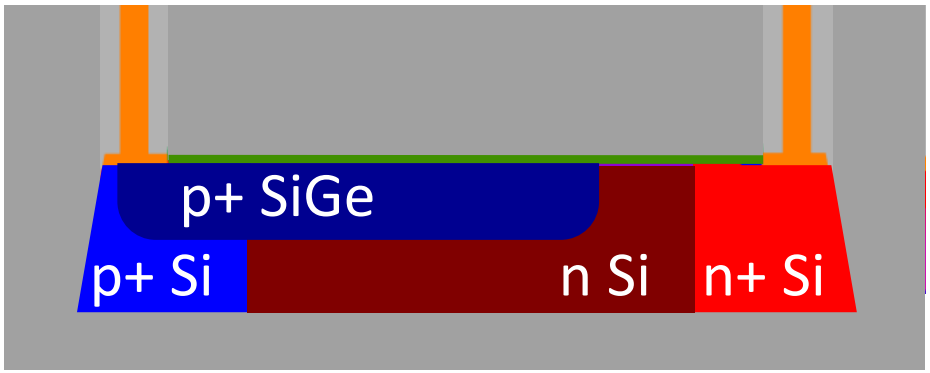
Depletion Modulator



Detector Integration

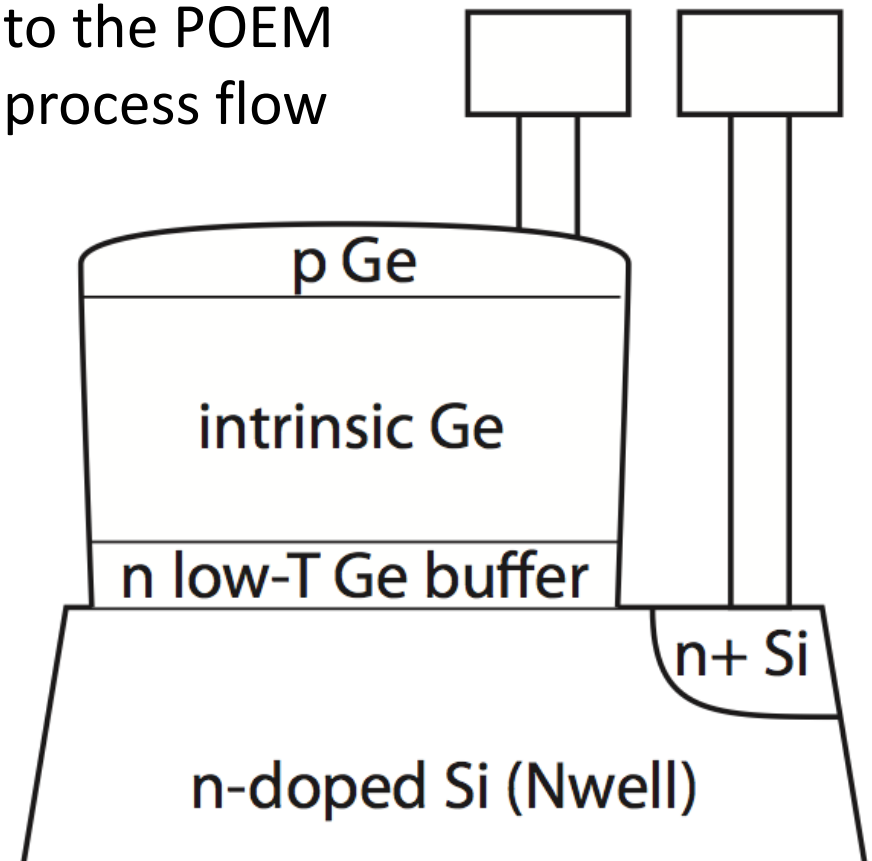


In the standard CMOS process low-mole fraction SiGe is available



Most DRAM and other memory processes do not include SiGe

Integrated Ge or SiGe added as a dedicated step to the POEM process flow



Front-End Process Compatibility Chart

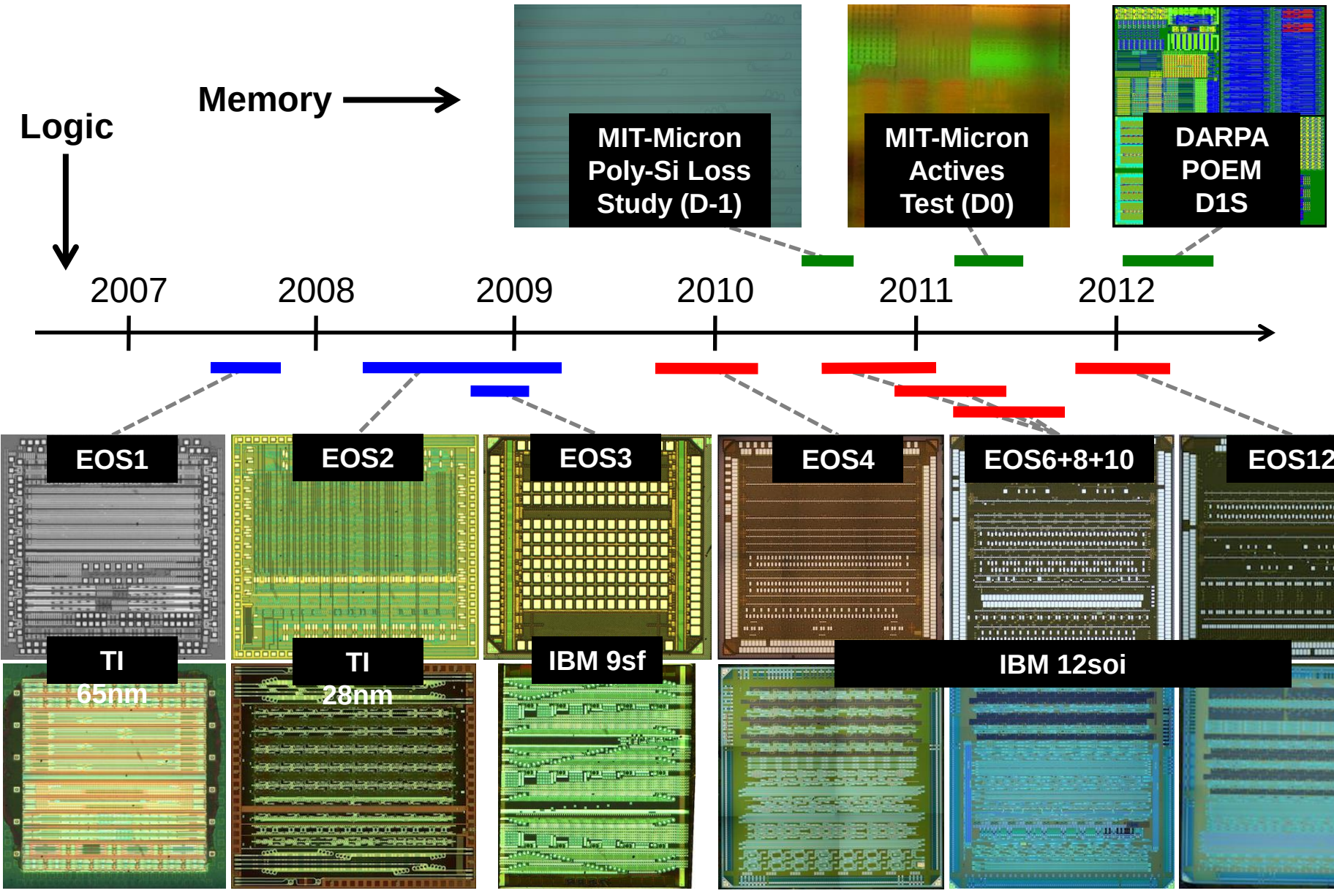
	CMOS (IBM 12SOI)	DRAM (Product)	DRAM (POEM)
Waveguide Core	**Silicon (3 dB/cm)**	**PolySi (~50 dB/cm)**	**PolySi (~6 dB/cm)**
Implant Levels	**Many**	**Few**	**Optimized**
Silicide Block	**Yes**	**No**	**Yes**
Mode Isolation	**Post-Process Only**	**Post-Process Only**	**Deep Trench**
Photodetector	**Doped SiGe S/D**	**2-Photon / Defect Si**	**Silicon Germanium**

Zero-Change Demonstrated

Not Possible In Most Cases

Optimized Process Required

Si Electronic-Photonic Integration Timeline



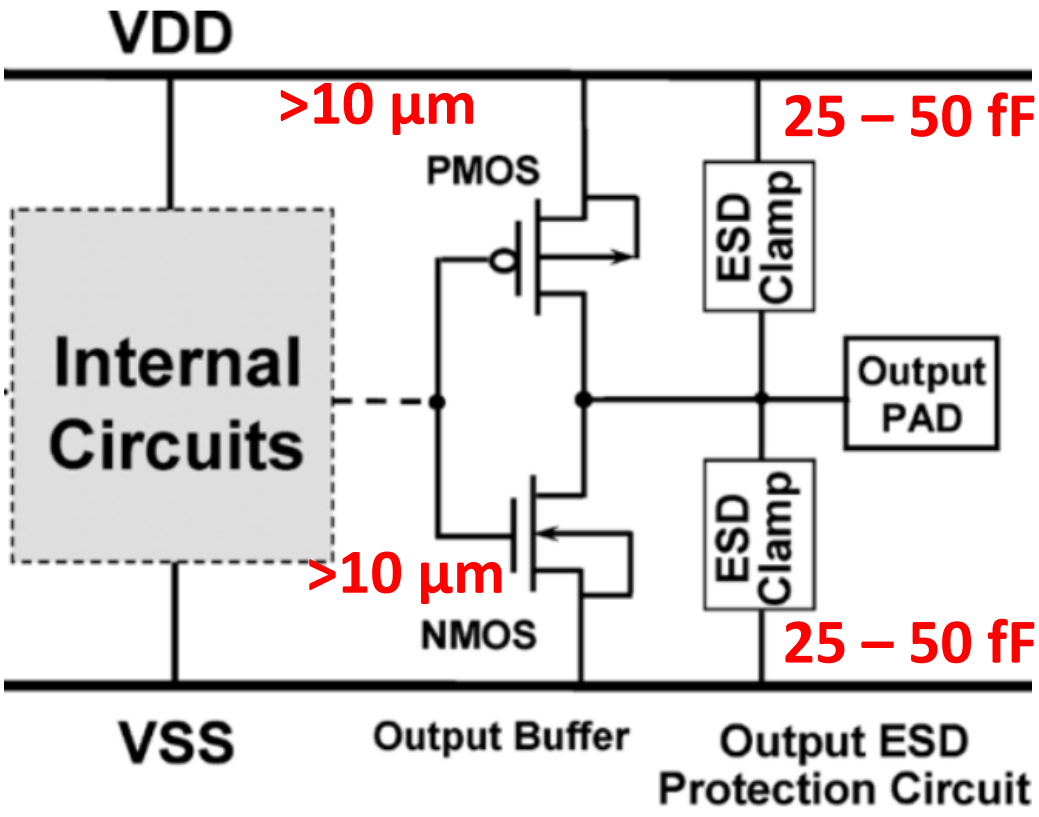
Monolithic Front-End Photonics

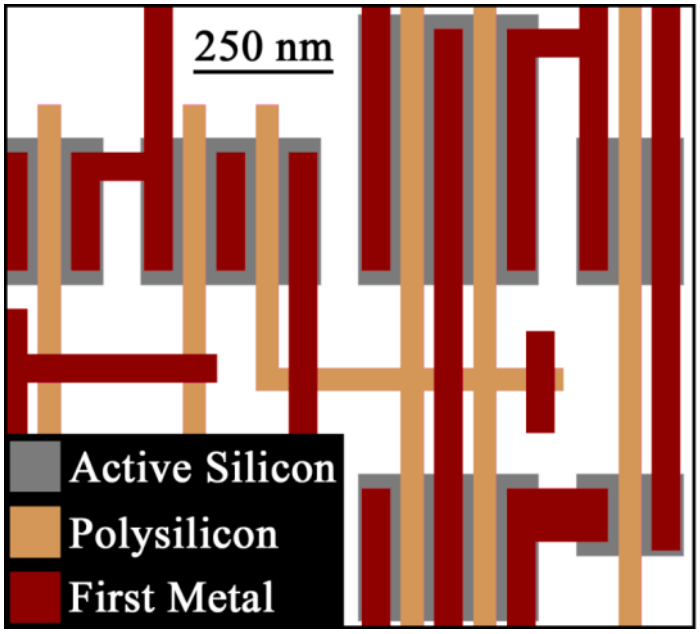
1. Process Integration
2. Area
3. Energy Efficiency
4. Alternative Schemes

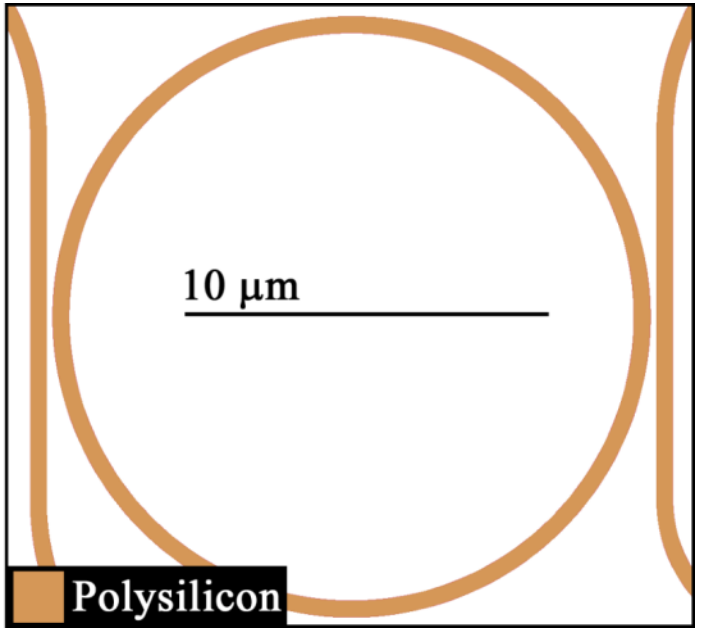
Are the integrated photonics too big?

Logic transistors are indeed dense and much smaller than the photonic devices

The I/O devices are much larger and must also include electrostatic discharge devices



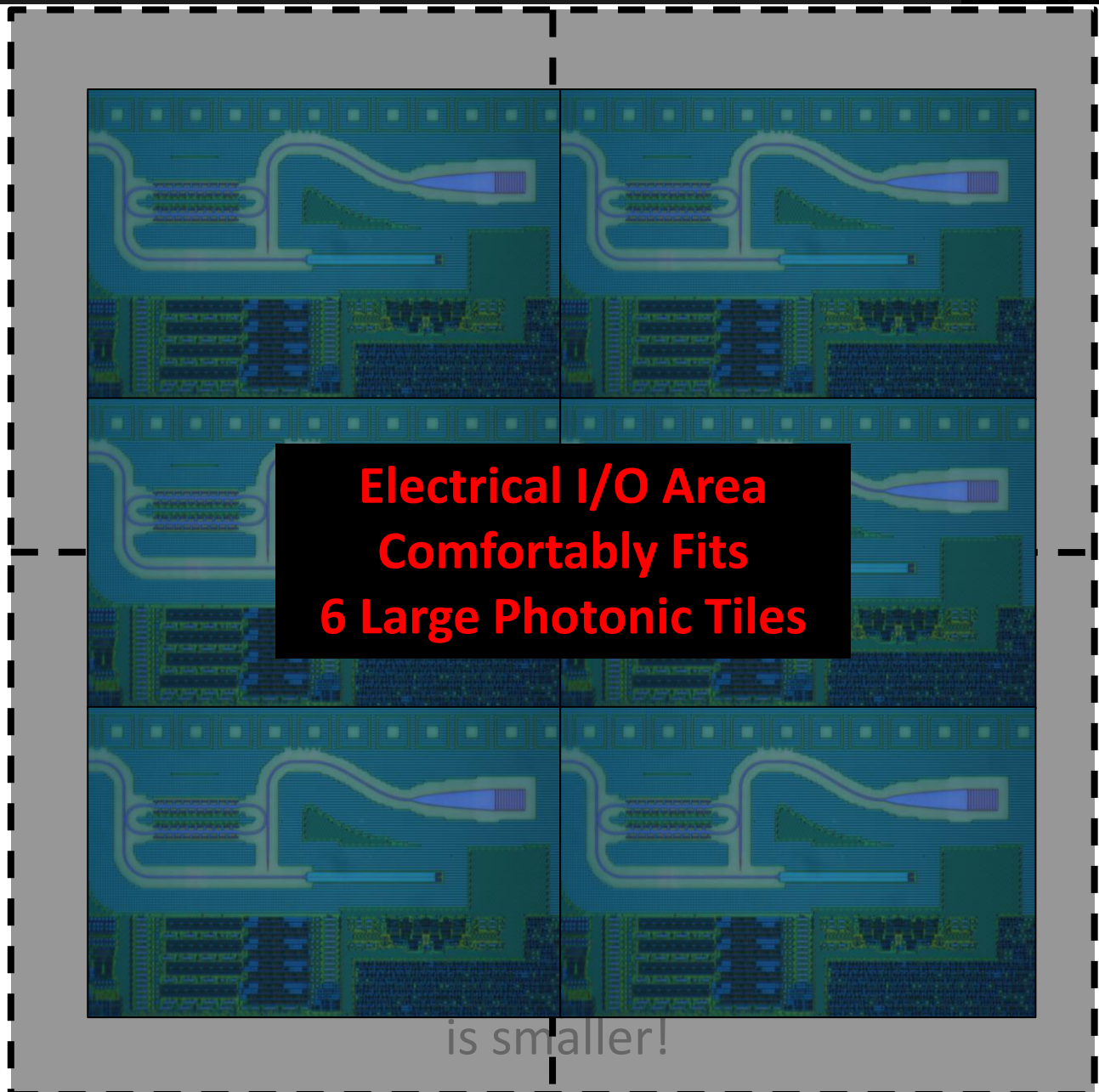




Area: Comparing Apples-to-Apples

The relevant area is what the photonics is replacing on-chip

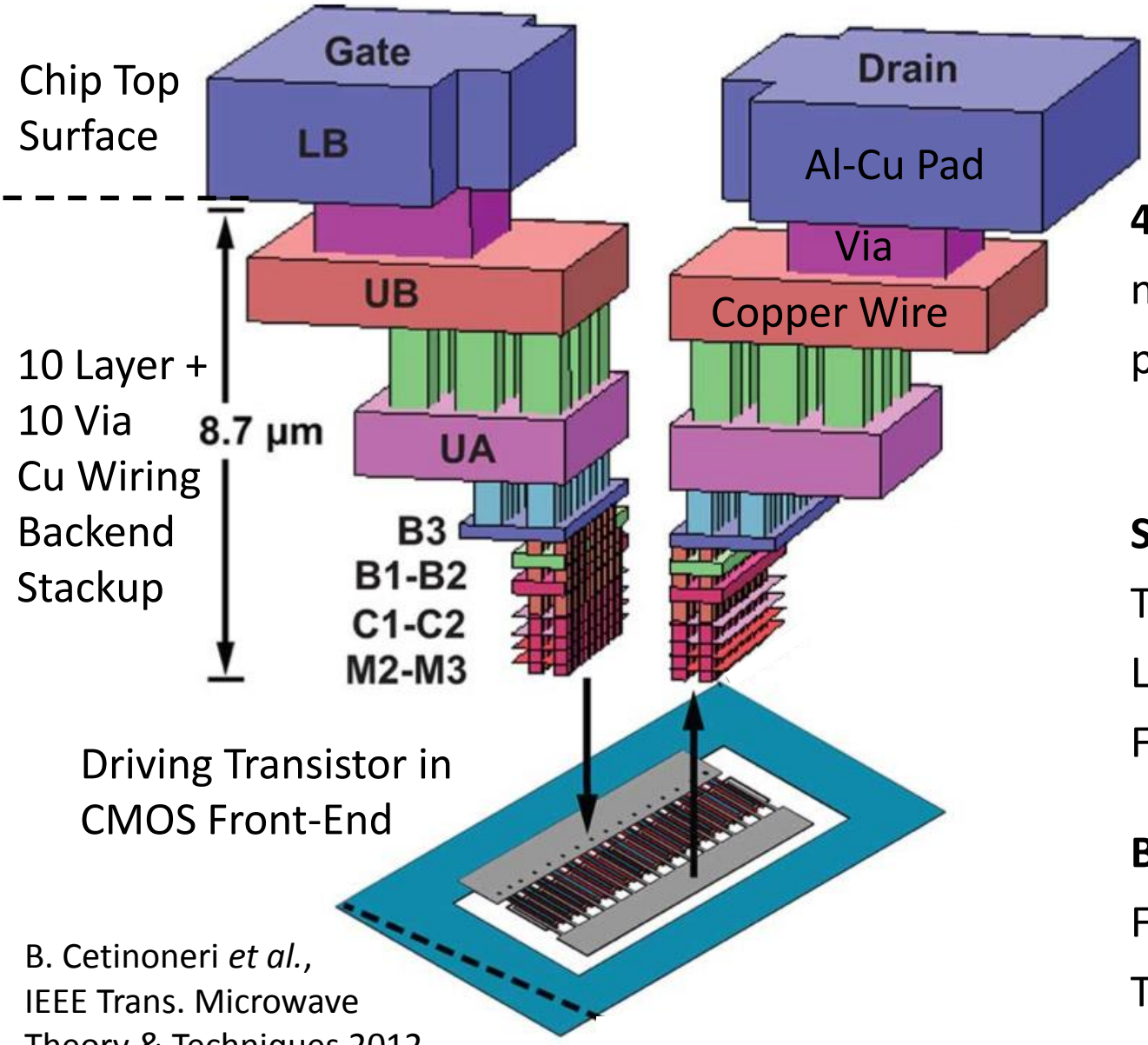
High-speed I/O limited by area array pitch and package fan-out



Monolithic Front-End Photonics

1. Process Integration
2. Area
3. Energy Efficiency
4. Alternative Schemes

Parasitics and Bandwidth in CMOS



B. Cetinoneri *et al.*, IEEE Trans. Microwave Theory & Techniques 2012

$$\text{“Speed”} \approx \frac{\text{Drive Current}}{\text{Capacitance}}$$

45nm SOI Drive Current

n-FET $I_{d,sat}$ = 1.24 mA/µm

p-FET $I_{d,sat}$ = 0.84 mA/µm

(S. Narasimha *et al.*, IEDM 2007)

Scaled CMOS Capacitances

Transistors = 0.3 fF/µm

Low Wires = 0.2 fF/µm

Full Via Stack ≈ 5 fF

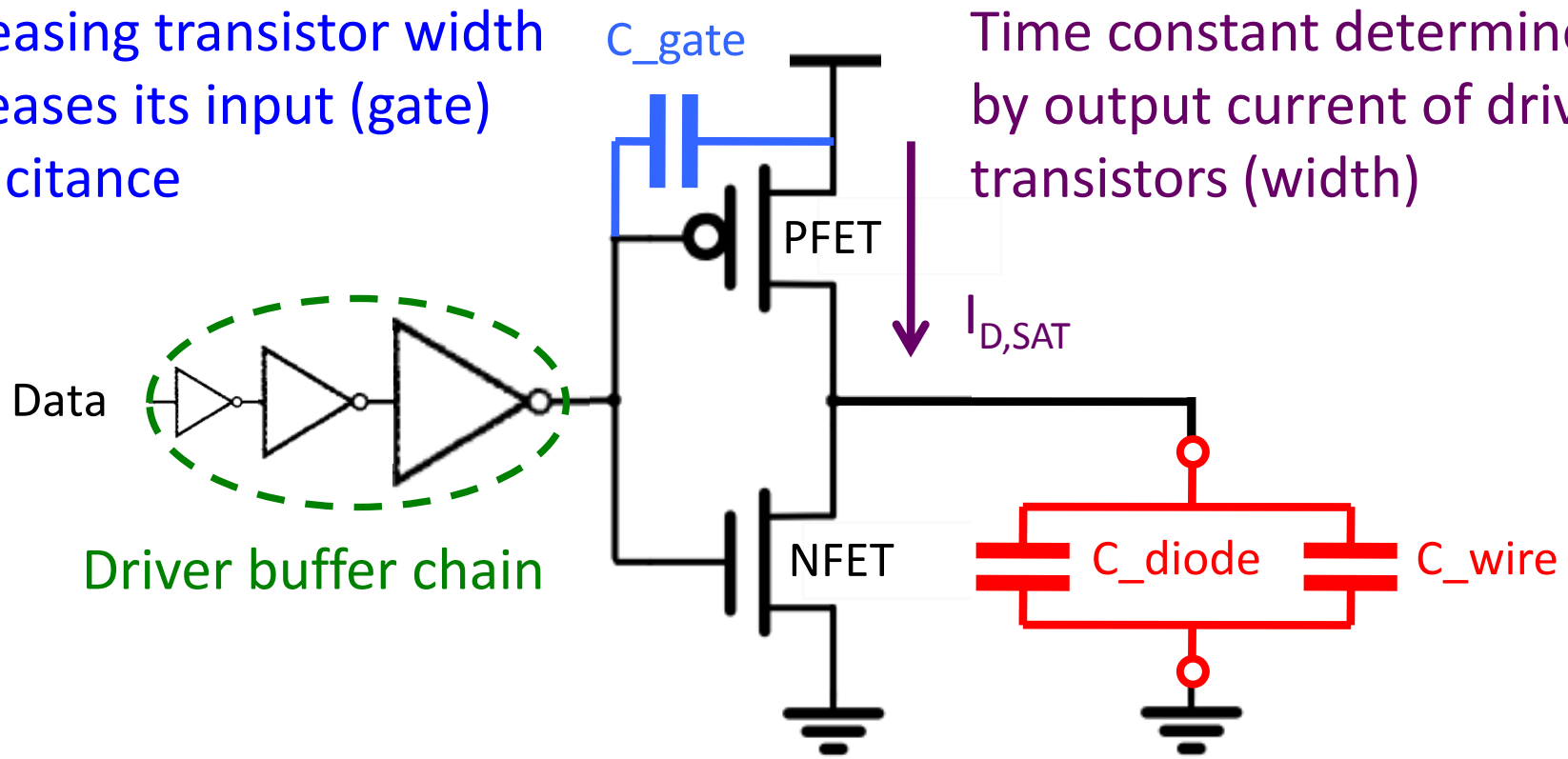
Bandwidths (45nm NFET)

First Metal = 340 GHz

Top Metal = 200 GHz

Modulator Energy Efficiency

Increasing transistor width increases its input (gate) capacitance



Time constant determined by output current of driver transistors (width)

$I_{D,SAT}$

As total node capacitance increases relative to the transistor stage delay for a given bit rate, total energy increases super-linearly

In depletion-mode operation, diode appears as a capacitor

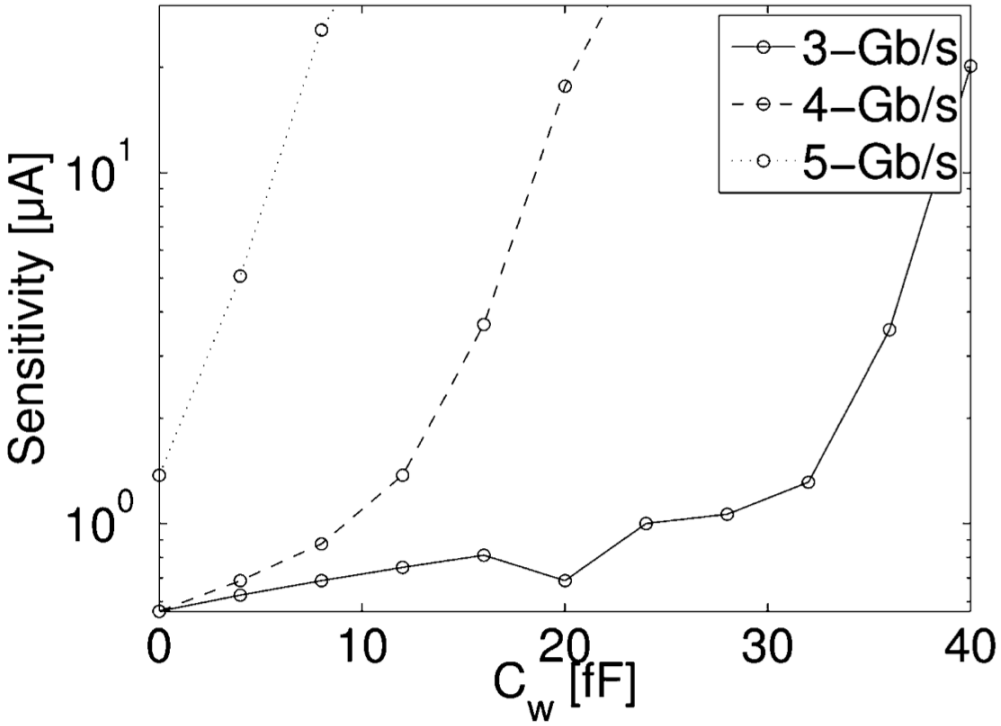
Energy per bit = ¼ $C_{tot}V^2$

Receiver Energy Efficiency

Monolithic integration enables direct detection receivers by low parasitic capacitances

$$V_{sig} = \frac{Q}{C} = \frac{\tau_{bit} I_{photo}}{C_{PD} + C_{wire}}$$

Capacitance and photocurrent trade equally for signal!



52 fJ/bit @ 3.5 Gbps with 4 μA sensitivity

M. Georgas, J. Orcutt, V. Stojanović
J. Solid State Circuits, July 2012

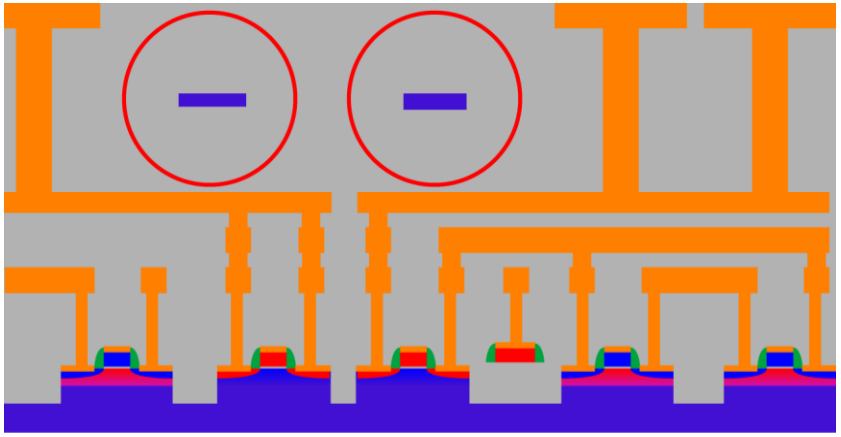
Monolithic Front-End Photonics

1. Process Integration
2. Area
3. Energy Efficiency
4. Alternative Schemes

Alternative Integration Options

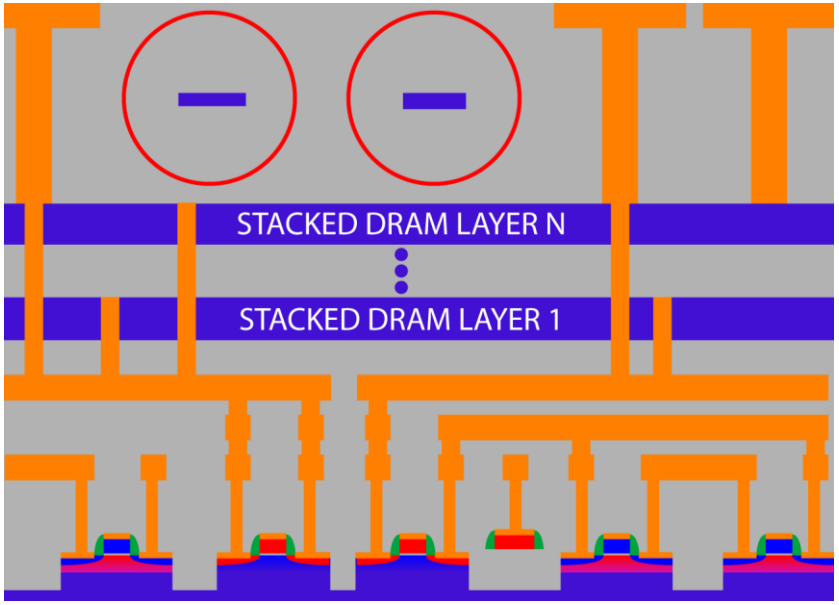
Backend Integration

- New process development
- New mask steps
- Low processing temperature

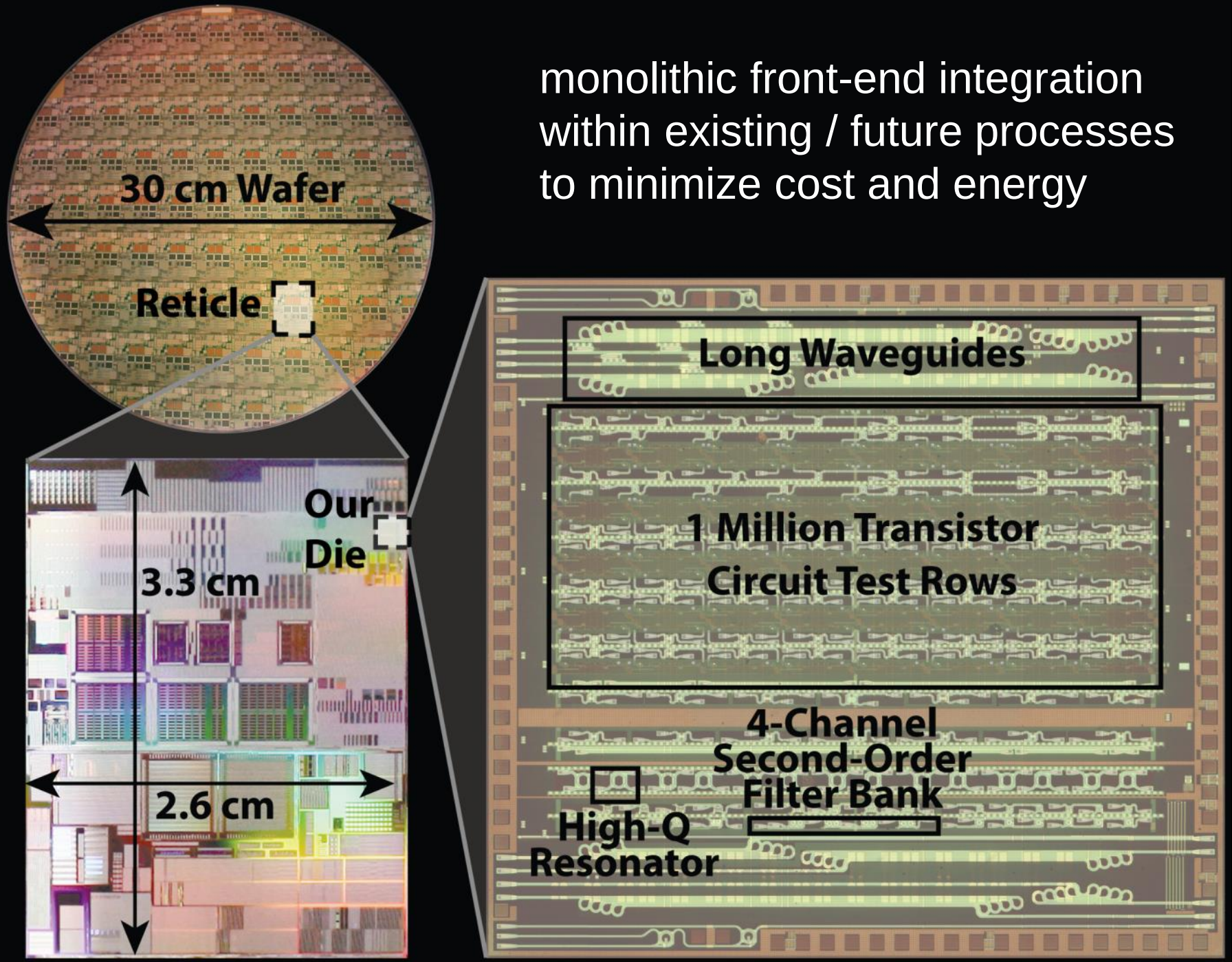


3D Integration

- High interconnection parasitics reduces energy efficiency
- Photonic layer still needs to fabricated



monolithic front-end integration
within existing / future processes
to minimize cost and energy



Conclusions

Process compliant integration enables monolithic photonics in scaled CMOS / DRAM products

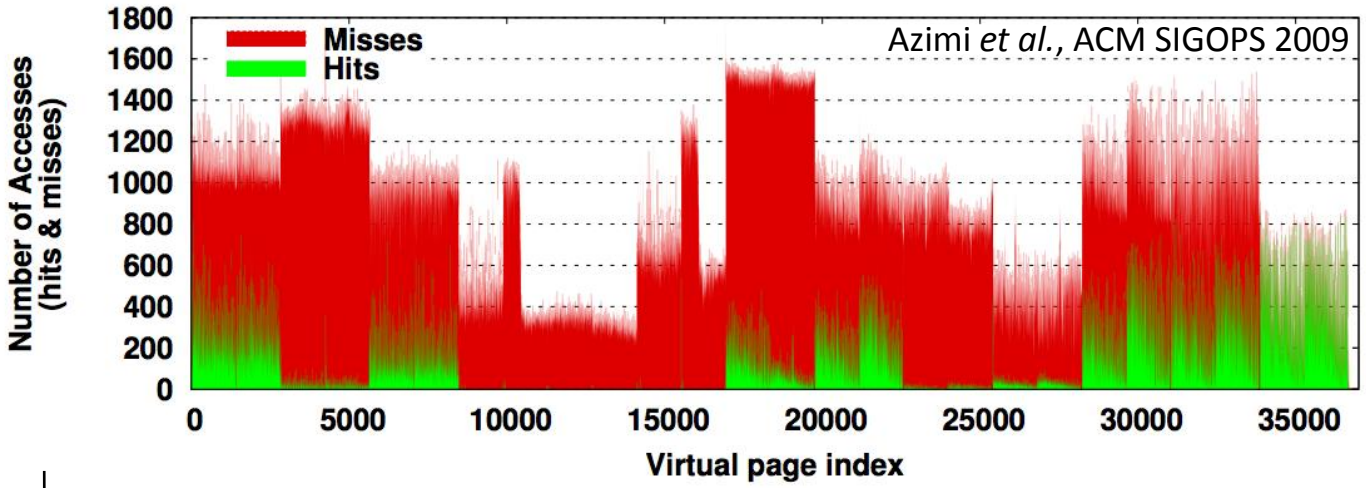
Monolithic photonics can provide an area **advantage** for microprocessors

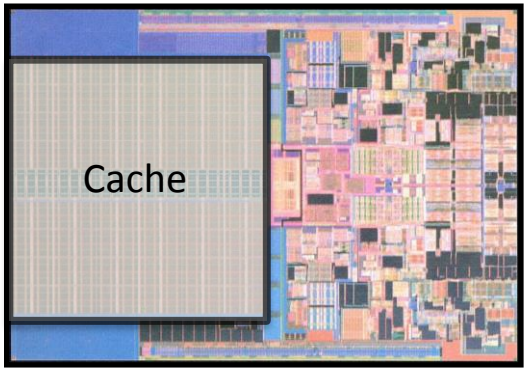
Energy efficiency may be optimized by front-end photonic integration

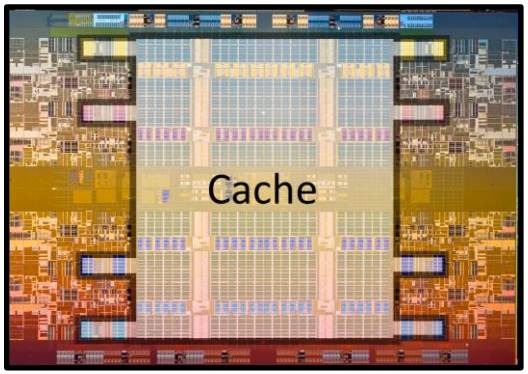
Area: Cache Reduction

Increasing memory bandwidth can eliminate the on-chip cache expansion

L3 Cache Miss Statistics for IBM Power5 running *swim-OMP*





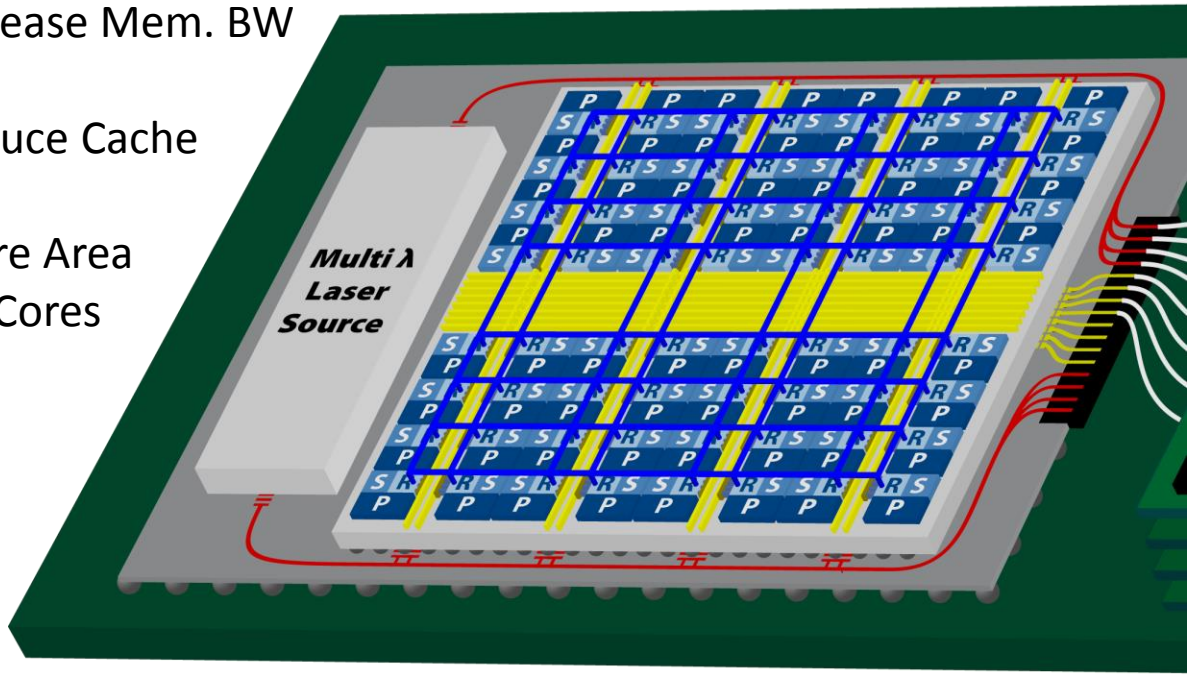


OR

Increase Mem. BW

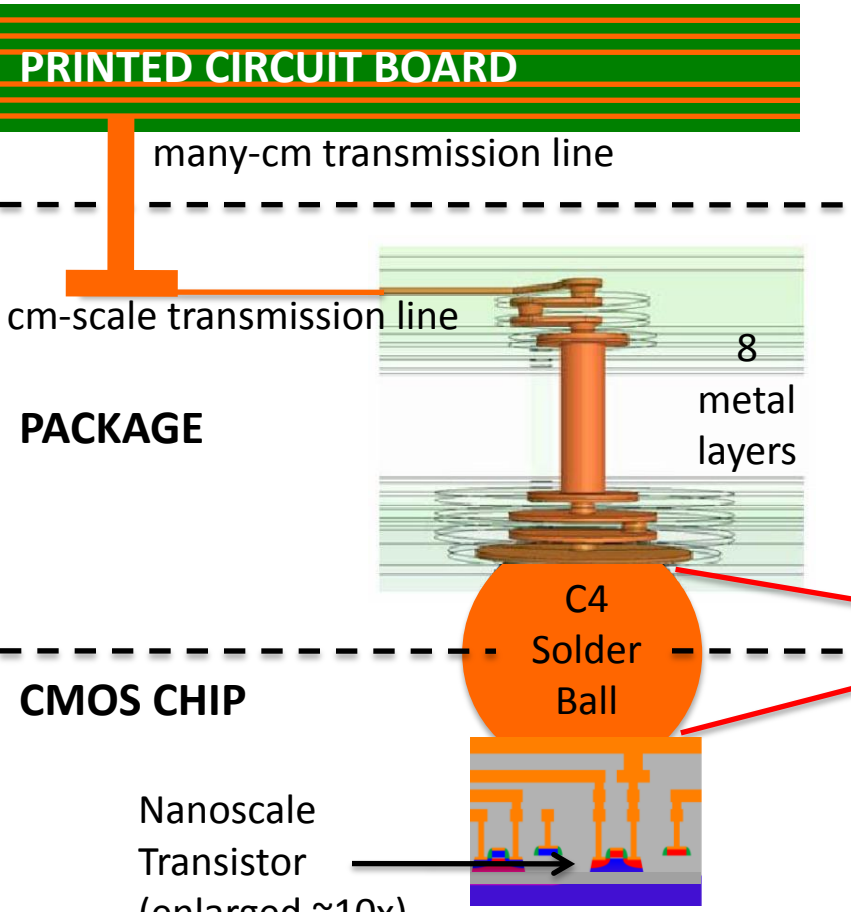
Reduce Cache

More Area for Cores



Off-Chip Link Does Not Scale Well…

Since 2006, Texas Instruments can build more digital signal processing cores on a single chip than they can feed with data



Off-chip electrical channel doesn't scale